Model Driven Paediatric European Digital Repository

Call identifier: FP7-ICT-2011-9 - **Grant agreement no**: 600932

Thematic Priority: ICT - ICT-2011.5.2: Virtual Physiological Human

Deliverable 17.4

Test on the prototype for the case- and ontology-based retrieval service

Due date of delivery: 29-02-2016

Actual submission date:

Start of the project: 1st March 2013
Ending Date: 28th February 2017

Partner responsible for this deliverable: HES-SO

Version: 1.1

Dissemination Level: Public

Document Classification

Title	Beta Prototype ofKDD& Simulation platform
Deliverable	17.4
Reporting Period	
Authors	HES-SO
Work Package	WP17
Security	Public
Nature	Report
Keyword(s)	Case-based retrieval, evaluation

Document History

Name	Remark	Version	Date
MDP_D17.4_v1.1	First draft	1.1	29/01/2016
MDP_D17.4_v1.2	Reviewed version	1.2	03/02/2016

List of Contributors

Name	Affiliation
Emilie Pasche	HES-SO
Patrick Ruch	HES-SO
Marcello Chinali	OPBG

List of reviewers

Name	Affiliation
Harry Dimitropoulos (internal rewiew)	ATHENA
Bruno Dallapiccola	OPBG

Abbreviations

CBR	Case-Based Retrieval
GUI	Graphical User Interface
DoW	Description of Work
IR	Information Retrieval

Table of Contents

1. Introduction

Physicians, who are facing complex diseases treatments, show a great interest in finding cohorts of patients similar to their patients. Thus, they can observe the response of a particular treatment and learn about the outcomes at different points in time (i.e. the episodes of care). Thus, the collected information may help the physicians to make clinical decisions. As part of the MD-PAEDIGREE project, different services based on various modalities (i.e. structured data, narratives, etc.) have been developed in order to identify similar patients. The case-based retrieval (CBR) service is one of them. It aims to help physicians to find patients similar to previously seen patients, based on some clinical syntheses (i.e. unstructured textual data). As input, the physician submits a description of his patient's condition (e.g. a clinical report describing a particular episode of care or a few keywords). As output, he obtains a list of episodes of care, ranked by relevance.

The testing and validation of a service is of major importance when designing a medical informatics application to ensure its use in real conditions. Horsky et al. [1] reported that two out of five of electronic information systems are abandoned or failed to fulfil the expected requirements. Therefore, it is essential to perform an assessment along three dimensions: the usefulness of a medical system (i.e. to ensure it fits with the requirements of the end-users), its robustness (i.e. to ensure it will not entail negative consequences) and its facility of use (i.e. to ensure its acceptation and use). The testing and validation can take place at different moments.

Kushniruk [2] described two approaches: a linear approach and an iterative approach. The linear approach includes different types of assessment at each stage of the development (e.g. user interviews during the planning phase). The iterative approach is more flexible and better adapted to the rapid and exploratory development of a system. It relies on the rapid development of intermediate prototypes, refined at each cycle of assessment until a final system meets the desired goals. The design, development and deployment of the CBR service are based on this iterative approach: a first – basic – system was build, and showed to physicians. After collecting their feedback, a second version – proposing more advanced functionalities – was developed.

In this deliverable 17.4, we present the testing and validation of the case-based retrieval service. The testing and validation of the ontology-based retrieval service will be presented in deliverable 17.5. Prior to the presentation of the qualitative and quantitative evaluation of the CBR, we will propose a brief description of the CBR service to facilitate the reading.

2. Existing infrastructure

Two versions of the CBR service have been developed. In this section, we will shortly present them. More details can be found in deliverable 15.1 (delivered M18) for the first version of the CBR and in deliverable 15.3 (to appear M42) for the second version of the CBR.

2.1. First version of the case-based retrieval service

The first version of the CBR service is based on a preliminary dataset provided by GNúBILA to HES-SO. This dataset is formatted as a CSV file. It contains medical records of 25,742 patients of the OPBG hospital

(Osepdale Pediatrico Bambino Gesù), all treated for some cardiac pathologies. First, the data have been indexed using Apache Solr (version 4.4.0) with default statistical tuning, corresponding to an approximation of Okapi BM25 [3]. A Graphical User Interface (GUI) has been developed.

The GUI is composed of two parts: a query part and an output part. In the query part (Figure 1), the physician provides information about his patient. He can either upload the data from a file (XML format) or manually fill the fields in the form. The output part (Figure 2) shows the similar patients. Up to 100 similar cases are retrieved, and are returned ranked by relevance (i.e. the first patient is the most similar to the patient of the physician). For each similar case, the following information is provided:

- The gender of the similar patient: male, female or unknown;
- The age of the similar patient: months are indicated if the patient is younger than three years old;
- MeSH terms that have been automatically attributed to the similar patient;
- An extract of the discharge summary limited to 10 words;
- A similarity score represented by a five-star system, based on the similarity between the query and the clinical reports of the similar patient;
- A link to the PCDR patient file to get more information about the similar patient.

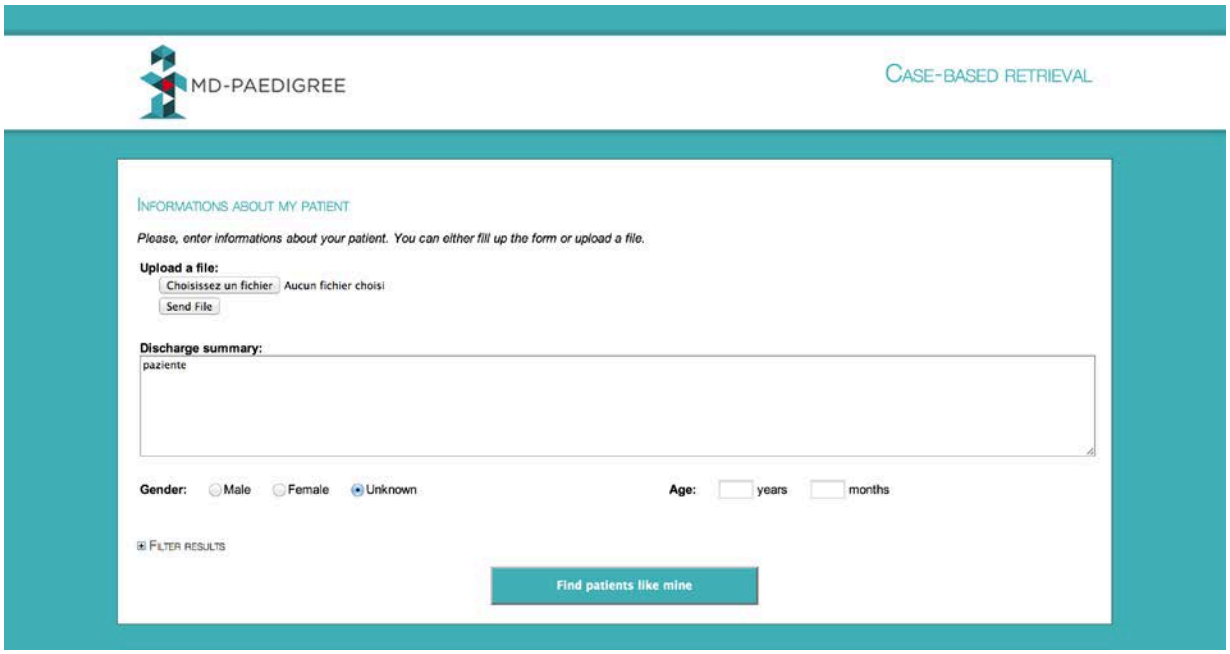
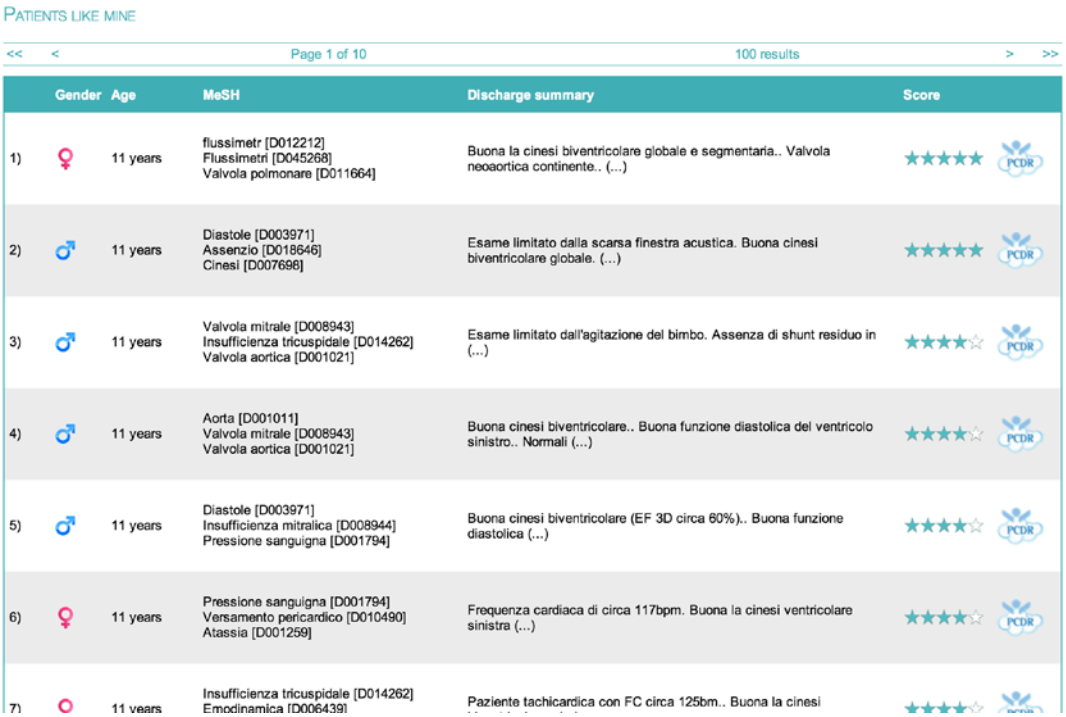


Figure 1 The query part of the first version of the CBR



PATIENTS LIKE MINE

<< < Page 1 of 10 100 results > >>

	Gender	Age	MeSH	Discharge summary	Score
1)	♀	11 years	flussimetr [D012212] Flussimetri [D045268] Valvola polmonare [D011664]	Buona la cinesi biventricolare globale e segmentaria.. Valvola neoaortica continente.. (...)	★★★★★
2)	♂	11 years	Diastole [D003971] Assenzio [D018646] Cinesi [D007698]	Esame limitato dalla scarsa finestra acustica. Buona cinesi biventricolare globale. (...)	★★★★★
3)	♂	11 years	Valvola mitrale [D008943] Insufficienza tricuspidale [D014262] Valvola aortica [D001021]	Esame limitato dall'agitazione del bimbo. Assenza di shunt residuo in (...)	★★★★☆
4)	♂	11 years	Aorta [D001011] Valvola mitrale [D008943] Valvola aortica [D001021]	Buona cinesi biventricolare.. Buona funzione diastolica del ventricolo sinistro.. Normali (...)	★★★★☆
5)	♂	11 years	Diastole [D003971] Insufficienza mitralica [D008944] Pressione sanguigna [D001794]	Buona cinesi biventricolare (EF 3D circa 60%).. Buona funzione diastolica (...)	★★★★☆
6)	♀	11 years	Pressione sanguigna [D001794] Versamento pericardico [D010490] Atassia [D001259]	Frequenza cardiaca di circa 117bpm. Buona la cinesi ventricolare sinistra (...)	★★★★☆
7)	○	11 years	Insufficienza tricuspidale [D014262] Emodinamica [D006439]	Paziente tachicardica con FC circa 125bm.. Buona la cinesi	★★★★☆

Figure 2 The output part of the first version of the CBR

2.2. Second version of the case-based retrieval service

The second version of the case-based retrieval service is based on a set of 47,433 episodes of care, corresponding to 33,674 distinct patients. The patients are consulting for cardiac pathologies. The source data originate from the OPBG hospital (Ospedale Pediatrico Bambino Gesù) and from the Taormina hospital. Data were obtained using the secured PCDR API developed by GNúBILA within WP14. The secured channel is the first step of the integration within the MD-Paedigree infostructure. The data have been indexed using Apache Solr (version 4.4.0) with a weighting shema tuned on a literature collection with similar distribution (average document length and average deviation). A Graphical User Interface (GUI) has been developed.

The GUI is composed of four parts: a query part, a refinement part, a filter part and an output part.

In the query part, the physician captures all information about the patient. There are currently two ways to provide this information. First, the clinician can type the patient identifier and the system will then load all historical clinical syntheses, as wel as demographic information (age and gender) for this patient (Figure 3). Second, he can manually fills the fields in the form so that ad hoc queries (published cases, cases extracted from cohorts, etc.) can be entered.

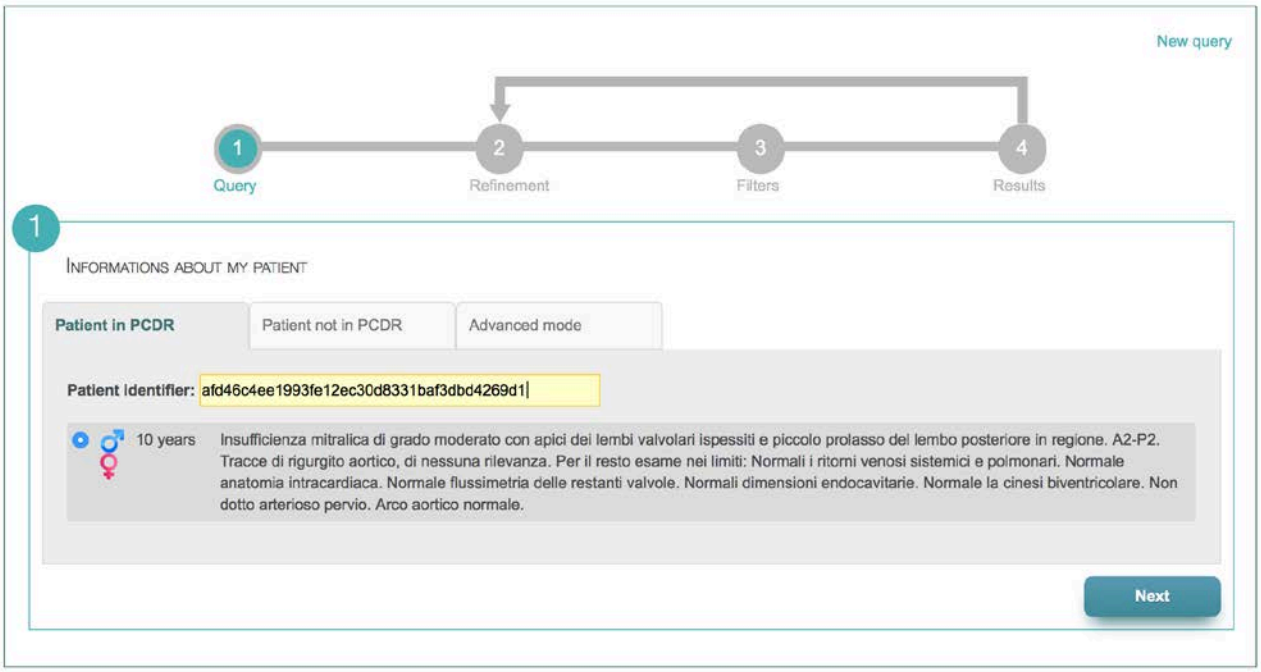


Figure 3 The query part of the second version of the CBR

The refinement part (Figure 4) proposes additional terms to be added to the query in order to – hopefully – improve the results' relevance. There are two query reformulation and refinement services proposed: a MeSH normalization of the query and a relevance-feedback (Rocchio) functionality. The MeSH normalization proposes up to 20 MeSH terms and the top-3 is by default pre-selected. The Rocchio component suggests additional keywords likely to be selected by the clinician. It is thus available only when the user has triggered a first search. The user can interactively select a few episodes of care he judges as similar or simply of interest to his patient and the Rocchio refinement services extract potential interesting keywords.

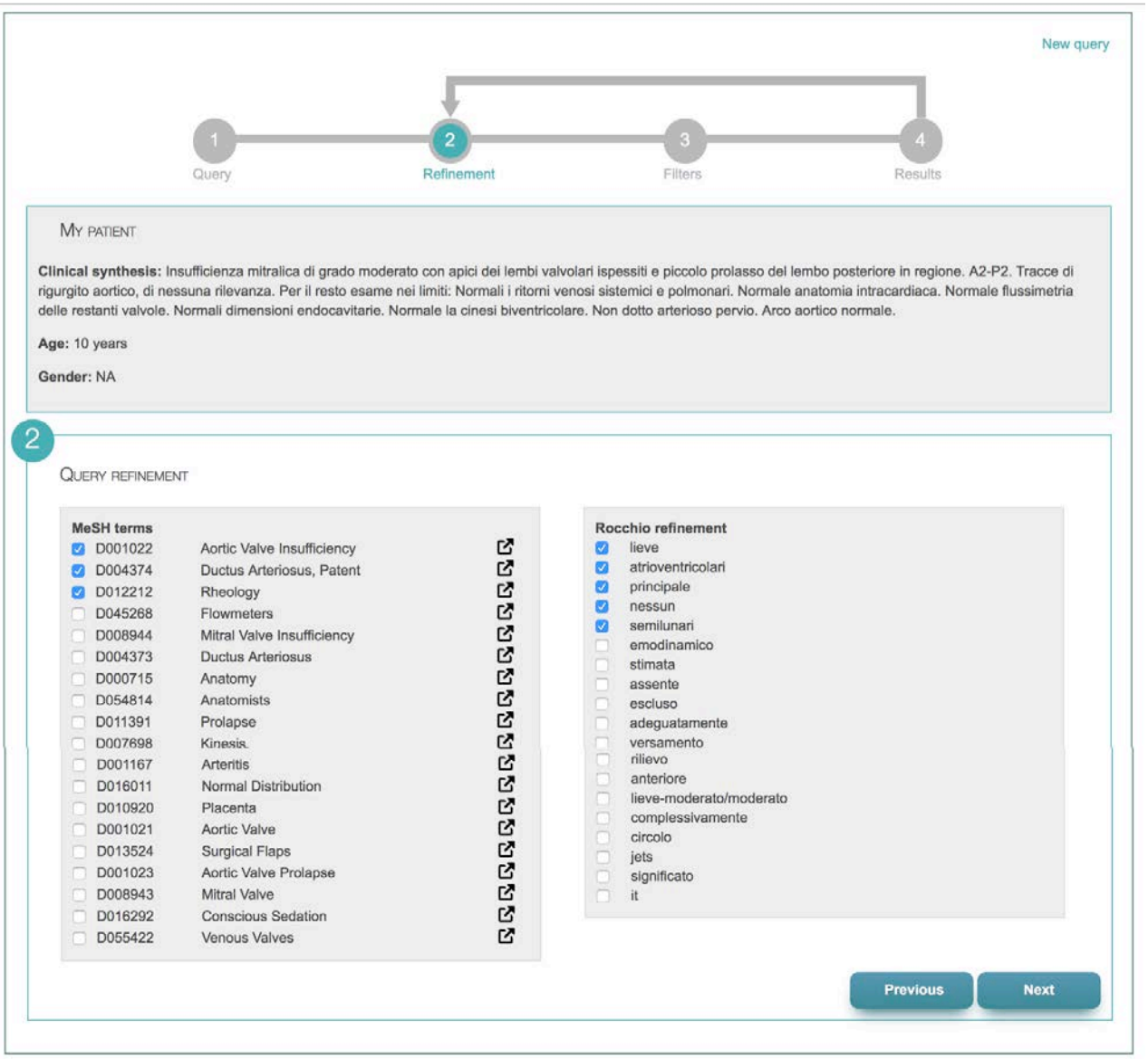


Figure 4 The refinement part of the second version of the CBR

The filter part (Figure 5) gives the opportunity to the user to modify his query before running it. Any element of the query can be removed. Additionnaly, the user can filter the output by age (e.g. show only patients from 3 to 10 years old) or gender (e.g. show only girls).

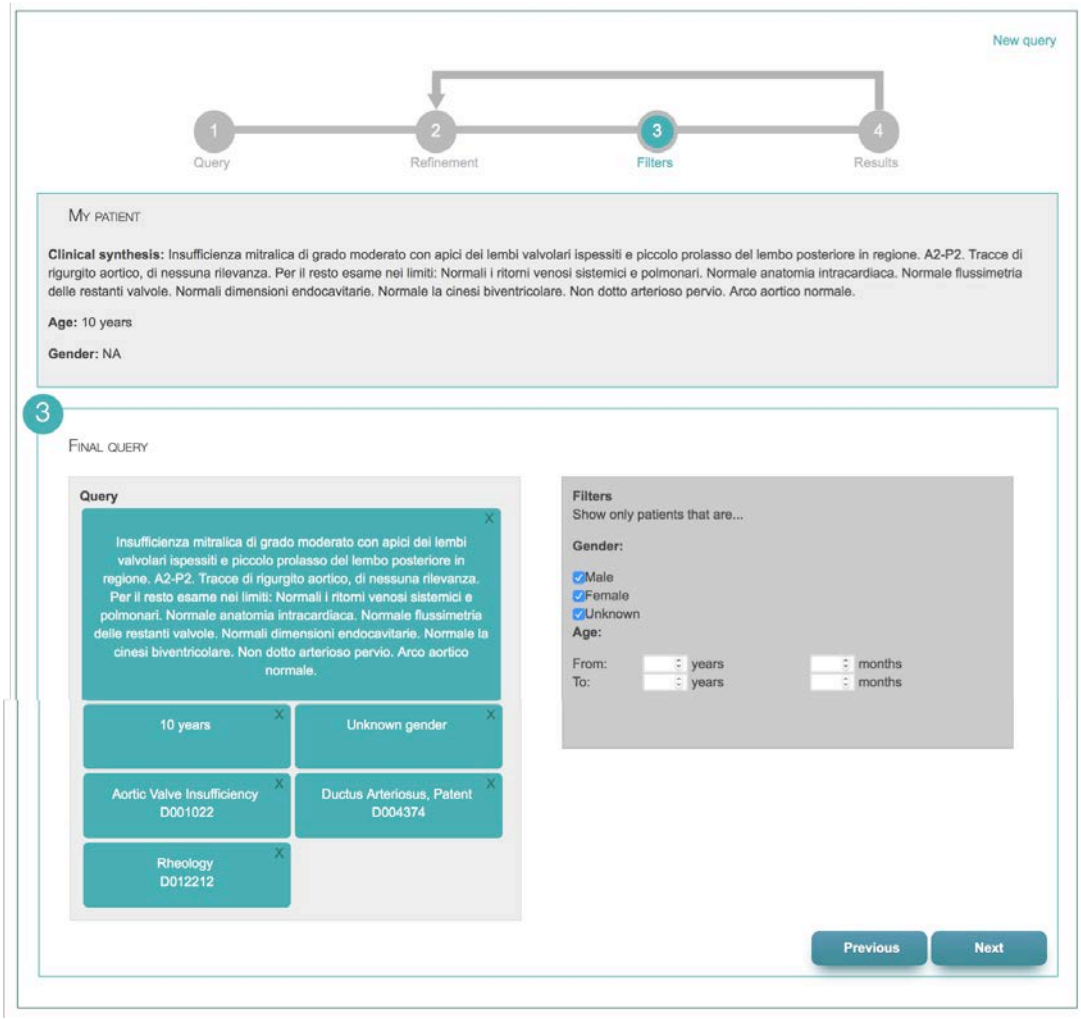


Figure 5 The filter part of the second version of the CBR

Finally, the display of the output (Figure 6) shows the similar episodes of care. Up to 100 similar episodes of care are retrieved and shown, ranked by relevance (i.e. the first episode of care is the most similar to the episode of care mentioned in the query). For each similar case, the following information is provided:

- The gender of the similar patient: male, female or unknown;
- The age of the similar patient: months are indicated if the patient is younger than three years old;
- MeSH descriptors, which have been automatically assigned to the similar episode of care;
- A summary is automatically generated out of the clinical syntheses of the similar episode of care. The physician can also access to the current episode of care's full clinical synthesis, as well as to the thread of all future clinical syntheses for the given patient just by clicking on the "Show similar and future events" button;
- The similarity score is represented using five-star icons;
- A link to the PCDR patient file to access all clinical information for each similar patient;
- Finally, a judgement panel represented by green and red smileys is available. The physician checks the green smiley if the episode of care is similar (i.e. relevant), and the red smiley if the episode of care is not similar (i.e. not relevant). This information is used for the Rocchio refinement as well as by the evaluation platform to benchmark the search effectiveness of the service.

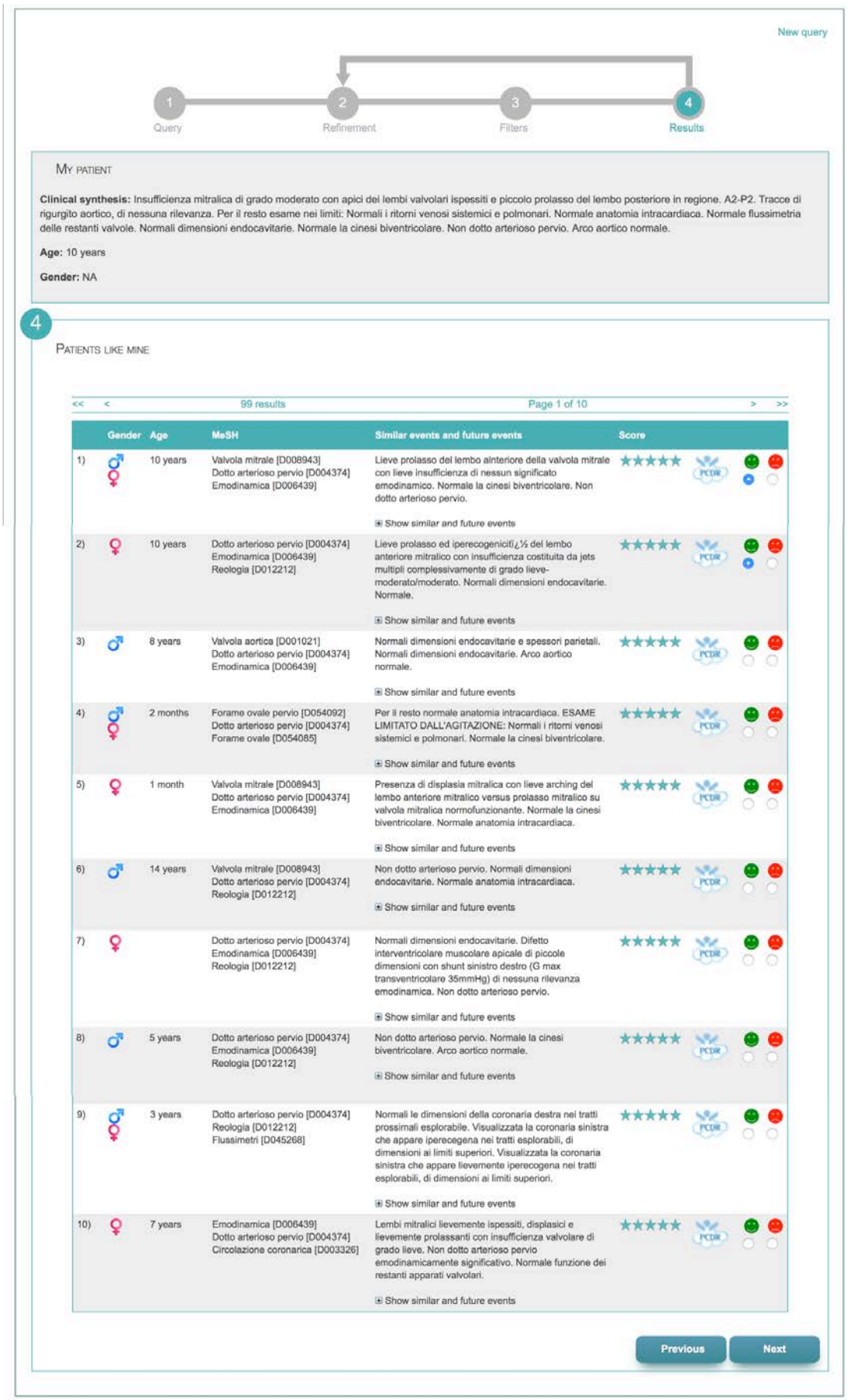


Figure 6 The output part of the second version of the CBR

3. Use case

3.1. First version of the case-based retrieval service

Name. Search for similar patients of a given patient.

Brief description. The physician enters the clinical report of his patient and obtains a set of similar patients.

Actors. A physician in paediatric cardiology.

Preconditions. The physician is allowed to access to PCDR patient files.

Post Conditions. The physician can make more informed decisions based on the outcomes of similar cases.

Basic Flow.

1. The physician describes his patient with free text and runs the system;
2. The system returns a ranked list of similar patients;
3. The physician browses the results, reads an extract of the clinical reports and finally accesses individual PCDR patient files to obtain additional clinical information;

Alternate Flow 1.

- The physician loads his patient's Electronic Health Records using a XML-formatted file and runs the system;
- *Steps 2-3 from the basic flow*

3.2. Second version of the case-based retrieval service

Name. Search for similar episodes of care of a given clinical synthesis.

Brief description. The physician enters the clinical report of his patient and obtains a set of similar episodes of care.

Actors. A physician in paediatric cardiology.

Preconditions. The physician is allowed to access to PCDR patient files.

Post Conditions. The physician can make more informed decisions based on the outcomes of similar cases.

Basic Flow.

1. The physician enters the PCDR identifier of his patient;
2. The system loads all clinical reports related to this patient;
3. The physician selects a clinical report and runs the system;
4. The system proposes a ranked list of MeSH terms that can be added to the query;
5. The physician selects/ignores the proposed MeSH terms and runs the system;
6. The system displays the final query and proposes a set of filters (age, sex);
7. The physician refines the filters if needed and runs the system;
8. The system returns a ranked list of similar episodes of care;
9. The physician browses the results, read a summary of the clinical reports or the full clinical reports of similar episodes of care and finally accesses individual PCDR patient files to obtain additional clinical information;
10. If needed, the physician can refine his query by selecting a set of episodes of care he considers as similar and iteratively runs the system again;
11. The system suggests additional keywords to add to the query based on a relevance feedback algorithm;
12. The physician selects the relevant keywords and runs the system;
13. *Steps 6-13 from the basic flow*

Alternate Flow 1.

- The physician describes his patient with free text and runs the system;
- *Steps 4-13 from the basic flow*

4. Evaluation

The two versions of the CBR have been assessed regarding two complementary dimensions: qualitatively (ergonomics, comprehensiveness of information…) and quantitatively (effectiveness, precision…). In this section, we will present the methodologies and results of the evaluations.

4.1. Qualitative evaluation

4.1.1. First version of the case-based retrieval service

The first version of the CBR has been evaluated during the third biannual meeting in Crete (October 2015). A demonstration of the tool has been conducted and the physicians of MD-PAEDIGREE provided their comments and recommendations. A synthesis of their remarks is presented below.

The main problem that arose from this demonstration session was the absolute demand to work at the episode of care level and not the patient level. Indeed, for the clinician, it is important to compare a patient at a given point of time and to follow the patient at different times to see the outcome. In this early version, indeed, all episodes of care of a given patient were merged together before indexing. Therefore, the systems generated search based on the full history of a patient, thus ignoring the clinical life cycle of the temporal dimension of healthcare.

The physicians were willing to access the entire clinical synthesis of the patient and not only a limited extract of the synthesis report (of about ten to twenty words) without navigating the PCDR. They also requested the possibility to easily observe the evolution of similar patients at different points in time. While this last aspect is possible using the direct link to the PCDR patient file, it assumed the physicians would go through all medical events, one by one, to finally access a particular clinical synthesis.

Regarding the query, some physicians expressed their interest in obtaining additional functionalities and in particular multilingual capacities. Indeed, for sake of demonstration all documents were in Italian, thus assuming queries should also be formulated in Italian. While the importance of such search capacities in a multilingual research environment is obvious, this request is questionable when targeting clinical decision support at the point of care. Additionally, the implementation of such functionality is relatively complex. The translation of clinical syntheses is a challenging objective: general-purpose automatic machine translation tools perform relatively poorly in highly specialised area such as healthcare [4]. An alternative option could be to leverage the availability of highly-specialized and normalized descriptors within the current infrastructure. Indeed, we automatically normalized the clinical syntheses with the MeSH terminology, which is available in virtually all European languages (e.g. Italian, English, German, French…). Therefore, a possible development would be to initiate a search using some MeSH descriptors.

Another set of suggestions addressed the possibility to weight the different parts of the query (i.e. some particular keywords). Indeed, the physicians are often searching using different keywords that do not all have the same importance for them, e.g. a primary diagnosis and a secondary diagnosis.

4.1.2. Second version of the case-based retrieval service

Following the demonstration session in Crete, a second version of the case-based retrieval service has been developed. Unlike the initial system, this new version was not patient-centric but episode-of-care-centric. It delivered a new search experience focused on the display of the fully set of clinical syntheses. In addition – and as planned in the DoW – a dedicated relevance feedback was implemented and integrated into the platform.

An evaluation session was held in Roma in January 2016. During this evaluation session, the two evaluators – MDs specialised in paediatric cardiology – searched for similar episodes of care using the second version of the CBR. A synthesis of the collected comments and recommendations is presented below.

In general, the evaluator appreciated the simplicity of use of the CBR service.

Nevertheless, a few technical problems have arisen during the evaluation. The two main problems observed were truncated reports and the failure to answer to some queries. The truncation was caused by a problem when importing the clinical syntheses in the MD-PAEDIGREE database from the OPBG IT system. Indeed, a limit to 255 characters was mistakenly set up in the ETL modules. Regarding the failure to answer some of the queries, it was due to inconsistencies in the generation of the Json exchange message in one of our APIs. All the identified technical problems were fixed right after the evaluation session.

The automatic MeSH normalisation triggered a strong interest from the audience, which is familiar with the terminological resources as it is used by the MEDLINE digital library – the legacy reference for healthcare litterature. For all the queries, the evaluators were ready to spend a few seconds to choose the appropriate MeSH descriptors. However, in some cases, it was noticed that the system did not suggest an existing relevant descriptor. For instance, for the query displayed in Figure 7, the evaluator was expecting a MeSH descriptor relative to the device (“device interventricolare”), which was missing. To improve its effectiveness, several suggestions have been made. First, the evaluator would like to have the possibility to manually add a MeSH descriptor in case the system skipped a potentially important one. Second, a stronger cleaning of the MeSH terms has been suggested (i.e. the term “Morinda” – a plant – in Figure 7 is not relevant in a cardiology context). Finally, the evaluators would appreciate to be able to select MeSH descriptor to be excluded from the results set: in the suggested similar episodes of care, the MeSH descriptor must not be present.

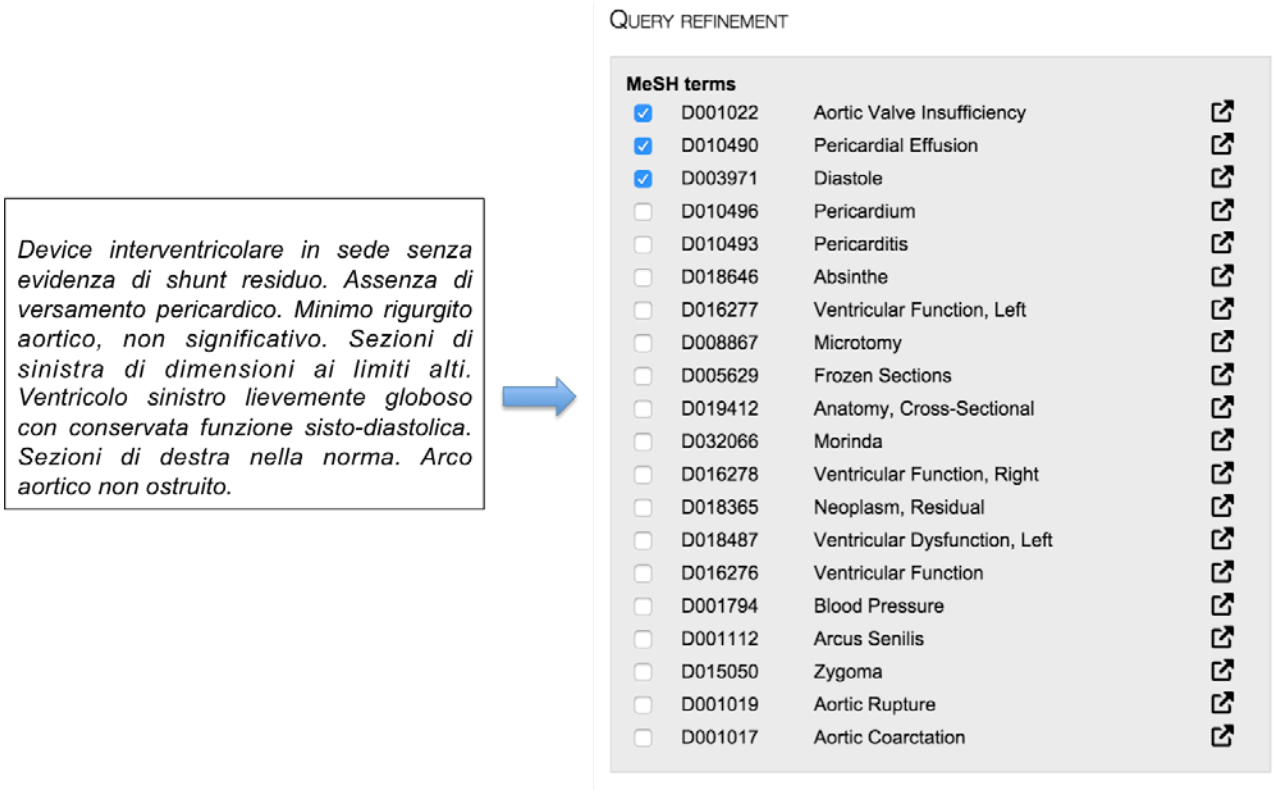


Figure 7 Example of a query

The Rocchio relevance feedback feature, which aims to suggest additional keywords to be interactively added to the query, showed some limitations during the evaluation session. The evaluators perceived the suggested terms as too general (i.e. common Italian words) or not clinically relevant. However, data analysis showed that for more than 90% of the queries, they selected a few terms. This feature is at a first stage of development and definitely needs to be improved. First, the list of terms should be cleaned in order to have only clinical and content-bearing terms. Second, the evaluators would appreciate the possibility to manually add a keyword. Third, the Rocchio algorithm could be improved to filter words with a high document frequency using IDF (Inverse Document Frequency), based on our collection. As an alternative to Rocchio, other feedback features are investigated, such as the latent semantic indexing in cooperation with UTBV.

Regarding the similar episodes of care suggested by the CBR, the evaluators reported that the system was very efficient to retrieve similar cases when the input case was a regular case. Dependent on statistical profiling frequent cases are simpler to handle. However, we report here some causes of failure (the system returned non-similar episodes of care in the top 10 documents). One major problem is the detection of the grade (e.g. normal, minor, severe, etc.). For instance, one of the queries described a patient suffering from a minor abnormality of the aortic flow, without structural abnormalities. All returned episodes of care were similar, except three of them, which had more severe abnormalities. An episode of care retrieved in position 6 was reporting a similar abnormality but with a stronger degree, while episodes of care retrieved in positions 3 and 10 reported cases with similar levels of abnormality but with a different prognosis (i.e a structural abnormality of the tricuspid valve).

Another remark concerned the length of the query. When the query is quite long, it happens that the similarity of the retrieved episodes of care is based on less important features than the primary diagnosis (e.g. secondary diagnosis, additional comments, etc.). In general, the evaluators reported better

performances with short and focused queries. In contrast, long and complex syntheses describe cases which occasionally are so rare that they are likely to be nearly unique. Another cause of confusion for the search engine is the occurrences of negation and doxic modalities (e.g. may be, unlikely…) in text, which are used to express nuances. Some episodes of care report that a patient is not or marginally suffering from a given diagnosis or sign. Such episodes of care are unfortunately retrieved by the system because they are lexically similar. Among the suggestions of the evaluators to solve these issues, the use of interactive facets to filter the results (e.g. all the retrieved episodes of care that report for a given diagnosis) would be an interesting feature.

The evaluators also tested the preliminary version of the Rocchio-based relevance service. The results were very diverse: for a few queries, some additional relevant documents were retrieved, for others unrelevant documents were added, while for some queries, the additional keywords did not bring any change in the ranking of the cases. To improve this service, the evaluators proposed that episodes of care judged not relevant during the first round should be discarded in the future iterations.

Another recommandation that was suggested is to offer the possibility to manually weight the elements of the query (e.g. in particular to increase the weight of the primary diagnosis or to decrease the weight of the age) in order to retrieve more relevant results.

4.2. Quantitative evaluation

Following the standard practice in the domain, pioneered by the Cranfield paradigm, [5], the quantitative evaluation of our search tasks is based on benchmarks. Benchmarks are constituted of three items: a corpus of documents, a set of queries and a set of relevance judgements. Because of the update of the data between version 1 and version 2, two different benchmarks have been created, following the same methodology. The corpus of documents is the set of patient files containing clinical syntheses, while the document unit changed between the two versions: 25,472 patient files for the first version and the 47,433 episodes of care for the second version. The set of queries consists in randomly selected clinical synthesis of 40 patients/episodes of care out of the whole corpus. The relevance judgements acquisition has been performed manually by experts in cardiology using the relevance judgement panel of the application. The top-10 results were manually checked and marked up using three categories: relevant (i.e. similar to the input case), irrelevant (i.e. judged as not similar to the query), or undecidable (only for the first version) if the information provided was not sufficient to determine the similarity.

The main evaluation in our settings is the precision of the search. Precision is the proportion of retrieved instances that are correct. In this evaluation, Precision at rank *i* (or P*i*) is the proportion of correct propositions in the first *i* ranks. Another common metric used in Information Retrieval is relative Recall, which is the proportion of correct instances in the collection that are retrieved. In IR evaluation, and in the MD-PAEDIGREE project, the assessor obviously did not inspect all the collection, for each case, in order to retrieve the comprehensive set of possibly correct answers, therefore the Recall is relative. Moreover, we can say that this task is somehow precision-oriented, i.e. the CBR engine does not aim at retrieving all the similar cases in the collection, but rather at retrieving some similar cases in order to extract useful information. Such an assumption will of course depend on the final usage but for decision-support it seems a valid hypothesis. Thus, we focused here on Precision at ranks 1, 5 and 10.

4.2.1. First version of the case-based retrieval service

The evaluation of the first version of the case-based retrieval service has been presented in deliverable 15.1. We present here a brief abstract of the results.

Out of the 400 cases, which were judged, the tag “yes” was assigned 219 times, the tag “no” was attributed 178 times and the tag “unclear” was attributed 3 times. The “unclear” tag was ultimately considered as a “no” judgement (i.e. not relevant).

For 8 queries, no similar case was found in the top-10. There are two hypotheses that might be considered to explain such phenomena: 1) the system was not able to find relevant documents for these queries; 2) the collection did not contain any relevant documents for these queries, meaning the case is so rare that there is no similar case. If the second explanation is valid then such queries are artificially decreasing the precision of the search engine. In the following, we separate the results into two sets: results based on the full set of queries (including the 8 queries with no relevant document identified) and results computed on the limited set of queries (excluding the 8 queries). The real precision of the system is therefore located between these lower and upper boundaries.

Table 1 shows different measures of Precision, for all queries, and for queries with at least a relevant identified answer. We display macro-average precisions: it means that precisions were computed by taking the average of the precision for each topic. The measured precisions are quite good: between two thirds and three quarters of the cases, the system is able to suggest a similar patient at the first rank. More than half of the top-10 patients suggested by the system are considered as similar to the patient of the query.

Parameter	All queries (40)	Queries with at least a relevant case (32)
P0	0.63	0.78
P5	0.59	0.73
P10	0.55	0.68

Table 1 Evaluation of the first version of the CBR engine.

4.2.2. Second of the case-based retrieval service

We present here the results of the evaluation of the second version of the CBR engine. Out of the 425 cases analyzed, the tag “yes” was attributed 188 times, the tag “no” was attributed 237 times.

Again, for 8 queries, no similar case was found among the top-10. The same hypotheses are applied to these data. For 2 queries, due to technical failure, no evaluation was performed.

Table 2 shows different measures of Precision, for all queries, and for queries with at least a relevant identified answer. In more than half of the cases and for up to two thirds of them, the system is able to suggest a similar episode of care at first rank. The observed precisions are a bit lower than for the first version of the CBR. However, the dataset is larger and the task is more challenging: to find a similar episode of care and not just a related patient.

Further, table 3 presents the results obtained with the relevance feedback algorithm. We observe a slight improvement of the P5 and P10 with the Rocchio-based results, which shown an improvement of the recall with a stable top-precision. Despite its very basic tuning at the moment of the evaluation (e.g. the evaluators reported the limited quality of the proposed keywords), we can thus consider that the gain brought by relevance feedback is worth being further explored.

Parameter	All queries (38)	Queries with at least a relevant case (30)
P0	0.5	0.63
P5	0.44	0.55
P10	0.42	0.54

Table 2 Evaluation of the second version of the CBR engine

Parameter	All queries (24)	Queries with at least a relevant case (19)
P0	0.5	0.63
P5	0.52	0.65
P10	0.45	0.56

Table 3 Evaluation of the Rocchio-based results of second version of the CBR engine

5. Conclusion

A methodology to develop and monitor the progree of the Case-base retrieval prototype has been implemented and tested. The initial and interim results were sufficient to improve the application regarding usability. From a quantitative point of view, the current results are alreay regarded as fair to support a case-based retrieval application, although several components, such as the relevance feedback service, needs fine-tuning to convince the end-users.

6. References

[1] J. Horsky, K. McColgan, J. E. Pang, A. J. Melnikas, J. A. Linder, J. L. Schnipper and B. Middleton. Complementary methods of system usability evaluation: surveys and observations during software design and development cycles. J Biomed Inform, vol. 43, no. 5, pages 782–790, Oct 2010.

[2] A. Kushniruk. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med, vol. 32, no. 3, pages 141–149, May 2002.

[3] S.E. Robertson, S. Walker, M. Beaulieu. Experimentation as a way of life: Okapi at TREC. Information Processing & Management, vol 36 (1), pages 95-108, 2000.

[4] P. Ruch, I. Tbahriti, J. Gobeill, A.R. Aronson. Argumentative feedback: A linguistically-motivated term expansion for information retrieval. In Proceedings of the COLING/ACL, pages 675-682, July 2006.

[5] http://trec.nist.gov/. Retrieved: 29/01/2016